Conditional Gradients for the Approximately Vanishing Ideal

Elias Wirth[1,2] Sebastian Pokutta[1,2]

[1]Technische Universität Berlin

[2]Zuse Institute Berlin

February 17, 2022

Problem Setting

- Input space $\mathcal{X} \subseteq [-1, 1]^n$ and output space $\mathcal{Y} = \{-1, 1\}$
- Training sample

$$S = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\} \in (\mathcal{X} \times \mathcal{Y})^m$$

  drawn i.i.d. from some unknown distribution $\mathcal{D}$
- Determine a *hypothesis* $h \colon \mathcal{X} \to \mathcal{Y}$ with small *generalization error*

$$\mathbb{P}_{(\mathbf{x},y)\backsim\mathcal{D}}[h(\mathbf{x}) \neq y]$$

Vanishing Ideal

- Set of feature vectors corresponding to class 1

$$X^1 := \{\mathbf{x} \mid (\mathbf{x}, y) \in S, y = 1\} \subseteq [-1, 1]^n$$

- *Vanishing ideal*

$$\mathcal{I}_{X^1} = \{f \in \mathbb{R}[x_1, \ldots, x_n] \mid f(\mathbf{x}) = 0 \text{ for all } \mathbf{x} \in X^1\}$$

- There exist finitely many *generators* $g_1, \ldots, g_k \in \mathcal{I}_{X^1}$ with $k \in \mathbb{N}$ such that for any $f \in \mathcal{I}_{X^1}$, there exist $h_1, \ldots, h_k \in \mathbb{R}[x_1, \ldots, x_n]$ such that

$$f = \sum_{i=1}^{k} g_i h_i.$$

- Similar for class -1

Feature transformation for linear kernel SVM

- Training sample

$$S = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\} \in (\mathcal{X} \times \mathcal{Y})^m$$

- Set of feature vectors corresponding to class ± 1

$$X^{\pm 1} := \{\mathbf{x} \mid (\mathbf{x}, y) \in S, y = \pm 1\} \subseteq [-1, 1]^n$$

- Construct sets of generators $\mathcal{G}^{\pm 1}$ for vanishing ideals $\mathcal{I}_{X^{\pm 1}}$
- Feature transformation

$$\mathbf{x} \mapsto \tilde{\mathbf{x}} = \left(|\mathcal{G}^1(\mathbf{x})|, |\mathcal{G}^{-1}(\mathbf{x})|\right)^\intercal,$$

  where for $\mathcal{G} = \{g_1, \ldots, g_p\}$, $|\mathcal{G}(\mathbf{x})| := (|g_1(\mathbf{x})|, \ldots, |g_p(\mathbf{x})|)$
- Train SVM on

$$\tilde{S} = \{(\tilde{\mathbf{x}}_1, y_1), \ldots, (\tilde{\mathbf{x}}_m, y_m)\}$$

`CGAVI`

Related Methods:

- *Approximate Vanishing Ideal algorithm* (`AVI`)
- *Vanishing Component Analysis* (`VCA`)

Our algorithm:

- *Conditional Gradients Approximately Vanishing Ideal algorithm* (`CGAVI`) employing the *Pairwise Frank-Wolfe algorithm* (`PFW`) for generator construction

Contributions

- Generalization bounds
- Sparsity
- Blueprint
- Results of numerical experiments

References I

[1] David Cox, John Little, and Donal OShea. *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*. Springer Science & Business Media, 2013.

[2] Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval research logistics quarterly*, 3(1-2):95–110, 1956.

[3] Daniel Heldt, Martin Kreuzer, Sebastian Pokutta, and Hennie Poulisse. Approximate computation of zero-dimensional polynomial ideals. *Journal of Symbolic Computation*, 44(11):1566–1591, 2009.

[4] Simon Lacoste-Julien and Martin Jaggi. On the global linear convergence of Frank-Wolfe optimization variants. In *Advances in neural information processing systems*, pages 496–504, 2015.

[5] Evgeny S Levitin and Boris T Polyak. Constrained minimization methods. *USSR Computational mathematics and mathematical physics*, 6(5):1–50, 1966.

References II

[6] Roi Livni, David Lehavi, Sagi Schein, Hila Nachliely, Shai Shalev-Shwartz, and Amir Globerson. Vanishing component analysis. In *International Conference on Machine Learning*, pages 597–605, 2013.